DOCUMENT RESUME

ED 174 676 TM 009 648

AUTHOR Cluxton, Sue Ellen; Mandeville, Garrett K.
TITLE Latent Trait Models: Ability Estimates and Omitted Items.
PUB DATE [Apr 79]
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Guessing (Tests); Item Sampling; *Mathematical Models; Primary Education; Reading Tests; *Scoring Formulas
IDENTIFIERS Birnbaum 2 Parameter Logistic Model; Latent Trait Model; *Missing Data; Rasch Model

ABSTRACT

A comparison was made between four different scoring procedures for the 45-item Reading Comprehension subtest, Level I, of the Comprehensive Test of Basic Skills, Form S, for a sample of 1,000 third grade students. These students were selected to have been among those who omitted from 3 to 22 of the 45 items. Another representative sample of 1,300 students was used to obtain one and two parameter estimates of the test items. The four scoring procedures included two procedures based on the Rasch (one-parameter) model and two procedures based on a two-parameter model. For each of these, one procedure was based on all 45 items, whereas the other procedure was based on only those items which the student had attempted. In general, the students received higher ability estimates when scored on only the items they had attempted, rather than on all items. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174676

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

LATENT TRAIT MODELS: ABILITY ESTIMATES AND OMITTED ITEMS

Sue Ellen Cluxton
Midlands Technical College and University of South Carolina

Garrett K. Mandeville
University of South Carolina

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sue Ellen Cluxton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper Presented at American Educational Research Association Meeting, San Francisco, 1979.

Session 32.21, Division D

TM009 648

INTRODUCTION AND OBJECTIVES

Standardized achievement tests invariably result in a substantial amount of missing data in the form of omitted items. Regardless of the number of omitted items, the examinee's score is the number correct. In some situations, it may be more desirable to employ a scoring rule which does not penalize an examinee who leaves a moderate number of items blank, say ten to fifty percent. Latent trait theory allows scoring rules to be developed which do not interpret omitted items as incorrect. Through the use of latent trait theory, it should be possible to obtain better ability estimates for such students; better here meaning that these estimates are more highly correlated with 'true' ability than are the raw scores. It is recognized that there exist some procedures for attacking this problem, but the issue here is one of practicality. It is felt that the method described here is more useable in practice. Both one-parameter (Rasch) and two-parameter models are investigated, thus providing a comparison of the utility of the procedure as a function of the number of item parameters in the model.

The major advantage which latent trait models have over classical test theory, is that given a set of items known to fit the model, a subject's ability may be estimated on the basis of any subset of items

that the subject attempts, i.e. the subject's ability is independent of the items which make up the test. The application of this advantage to the problem of missing data seems clear. The item parameters may be obtained from an analysis of the results of a large group administration of the test. With these known parameters, ability estimates of subjects answering a subset of the test can be obtained by considering only that subset. Although some computer programs, e.g. LOGIST (Wood and Lord), do differentiate omitted responses from incorrect responses, these programs require large amounts of computer time and core requirements for large data sets. The approach taken in this paper was to obtain estimates of the item parameters for the one parameter (Rasch) and two parameter logistic model for a large group of subjects, and then employ these to estimate the abilities of students omitting items. It was felt that if this procedure would provide estimates which were of similar validity as those obtained from the more extensive procedures, they might be more likely to be employed by the research community. It must be emphasized that this procedure to be discussed will not necessarily provide more valid ability estimates for everyone who omits items.

It is obvious that missing data presents a philosophical in addition to a statistical problem. Examinees vary in their propensity to take risks. Some will respond to all items, even if many responses are virtually random, while others, being more timid, will leave an item blank unless they are reasonably certain that their response is correct.

Others, particularly the slow readers and/or generally less capable students, due not complete the test due to time constraints. Test publishers such as CTB/McGraw-Hill, since their norms are developed on the basis of a fixed-length, k item test, count all such responses as incorrect. Although the logic of discounting omitted items throughout the body of the test is reasonably compelling, for students omitting the last r items of a k item test, it is reasonable to assume that they 'took' a k-r item test for which the k item norms do not apply. Hambleton and Cook (1971) take this position when they place the following qualifier on application of the one, two and three parameter logistic models:

> "All three functions can be applied to binary-scored items administered under non-speeded test conditions. (When the items are administered under speeded test conditions, it is necessary to distinguish between 'omitted items' and 'not reached items', so as to properly estimate examinee ability scores.)"

Scoring rules considered will employ the one parameter (Rasch), and two parameter logistic latent trait models. The commonly available algorithms for processing data using these models vary in the way missing data are construed. For example the Rasch program (BICAL) interprets omitted items as incorrect, whereas the three parameter program (LOGIST) places a different interpretation on blanks in the response vector. It

is the objective of this paper to investigate the properties of ability estimates derived from these algorithms, and a modified Rasch algorithm, for response vectors containing missing data.

PROCEDURES

In this work, two general approaches were taken. First, actual response data were submitted to various estimation algorithms and the parameters associated with the one and two parameter logistic models were estimated. This data consisted of both complete and incomplete responses, where incomplete responses could range from one item to many items. A question may arise here as to whether the population parameters ought to be derived from only complete responses, and a case may be argued for either stand. The position here was taken that the parameters estimated from the mixed group (as regards completeness of response vector) represented more realistically the population investigated.

The population parameters were obtained from BICAL (for the one parameter case) and from LOGIST (for the two parameter case). For BICAL (Wright and Mead, 1978) the maximum likelihood estimation option was taken. The computer program LOGIST (Wingersky and Lord, 1973) was used for the purposes of estimation under the two parameter logistic model. This program also provides maximum likelihood estimates and

handles omitted items internal to the exam by assigning a $1/A_i$ (A_i = No. possible response categories to item i) to the response in the likelihood function rather than a 0 or a 1. It also only considers the items a respondent 'completes' as far as the ability estimate is concerned.

These item parameters were then taken as population values for the purpose of application to subjects with incomplete response vectors. For this phase, students omitting ten to fifty percent of the items on the standardized achievement test were identified. Using the item parameters estimated from the large sample, as described in the data source, corrected and uncorrected ability estimates were obtained for these subjects. Uncorrected estimates are those treating missing items as incorrect. Corrected estimates were obtained by modifying the standard (one and two parameter) algorithms so that only ability estimates were obtained by iteration. The corrected estimates were than related both to the LOGIST three parameter estimates and the CTB raw scores. These data were processed by the Rasch procedure to provide uncorrected estimates; by a modified one parameter (Rasch) approach to provide one parameter corrected estimates; and by the LOGIST procedure for two parameters to provide corrected two parameter estimates. The modified Rasch approach, based on item difficulties from BICAL, produced ability estimates using an algorithm which did not attend to blank responses. Essentially, this iterative

algorithm mirrors the ability estimation phase of the one parameter model, i.e.

$$\hat{\theta}_1 = \ln\left(x/(n_a - x)\right)$$

where x = raw score and n_a = no. items attempted and

$$\hat{\theta}_r = \hat{\theta}_{r-1} + \left(x - \sum_g P_g\right) \Big/ \sum_g P_g Q_g$$

where

$$P_g = \frac{e^{(\hat{\theta}_{r-1} - b_g)}}{1 + e^{(\hat{\theta}_{r-1} - b_g)}}$$

and the summations above are over the n_a items attempted. The b_g above represent the Rasch parameter estimates. In addition, the LOGIST three parameter procedure was applied to this group to provide a 'true' ability estimate, the rationale being that this would represent the best ability estimate available, taking into account items not reached, items omitted, and guessing.

Thus from these procedures, four estimated abilities were available for each of the subjects.

The difference between the corrected and uncorrected ability estimates has been investigated as a function of number of omits, percentage of omits at the end of the test, average difficulty of omitted items, and percent of omitted items beyond the subject's estimated ability.

DATA SOURCE

The data consists of a portion of that collected by the 1978 South Carolina Statewide Testing Program. That portion discussed here comes from third graders who were administered Level I of the CTBS/S. The subtest providing the basic data is the 45 item Reading Comprehension subtest, selected because of its relative length, and the importance of the skill it tests.

The population group was a systematic one-in-twenty-five sample of the approximately 35,000 subjects, where only records with multiple responses to an item were eliminated. The sample size, truncated for simplicity, was 1300 subjects.

The sample of subjects with appropriate incomplete responses was a different systematic sample, of size 1000. A one-in-three systematic sample from the original 35,000 was screened for subjects omitting 3 to 22 of the 45 items, and from this group, a systematic one-in-six sample was selected and truncated to 1000. Roughly fifty percent of the records indicate 3 to 22 items omitted. The possible overlap between population group and incomplete sample, has been determined to be less than two percent.

RESULTS

As the correlation between the modified one parameter estimates and the two parameter estimates was quite high (.96), primary attention was directed towards relating the modified one parameter estima-es to the LOGIST three parameter estimates and to other variables of interest. The modified one parameter estimates are basically simple to obtain, and make few time and space demands on the computer. It is desired here to demonstrate their comparability to the LOGIST estimates.

The difference between the corrected (modified one parameter) estimates and the uncorrected (Rasch) estimates, hereafter referred to as the difference in ability estimates, can be quite large. Approximately seventy percent of the subjects obtain an increase of ability estimate of 0 - .7 logits, although this varies with whether or not the omitted items were at the end of the test or in the body of the test, as shown in Table 1. The greatest benefit is realized by those subjects omitting items only at the end of the test. This group, in large part, would represent those who lack the time to complete the test. The least benefit is realized by those who omit items mainly in the body of the test. This group would include those who omit items which they perceive as being too difficult as they progress through the test, and it is these items for

which the probability of a correct response would be small.

The difference in ability estimates was found to be significantly correlated with the number of items omitted, the tendency to omit difficult items (negative) and the tendency to omit items at the end of the test. These correlations are given in Table 2.

As a group, only about ten percent of the subjects omit items both at the end of the test and in the body of the test. The remaining ninety percent are approximately equally split between omitting only at the end, and omitting only in the body of the test. As regards difficulty, for one third, all omitted items are too difficult, and for one third, all omitted items are within their ability.

The modified one parameter procedure does produce results similar to the LOGIST results, as evidenced by correlations ranging from .96 to .98. These correlations, along with correlations for Rasch estimates, two parameter estimates, and raw scores are presented in Table 3. The variable which most strongly affects the correlation between the modified one parameter estimates and the LOGIST estimates is the number of omissions. As the number of omissions increases, this correlation decreases from .98 (3 - 7 omissions) to .87 (18 - 22 omissions), as is shown in Table 4.

A scatterdiagram of the 'true' ability and the modified one parameter ability is presented in Figure 1, where it may be seen that the divergence between the modified one parameter estimates and the

LOGIST estimates is greatest for high ability estimates. These are subjects who respond correctly to all but one or two of the items they attempt. As with LOGIST, the modified algorithm, in the circumstance of a subject correctly answering all attempted items, will result in an unrealistically high ability estimate. There were three such subjects out of the sample of 1000. Their abilities were limited to the maximum of the range of all other subjects.

CONCLUSIONS

The focus here has been on alternative methods of scoring fixed length (k item) tests containing missing data. The overriding objective has been to identify methods of scoring such tests which do not overtly penalize students who do not complete the test. It was also hoped that any method identified as useful in this regard could be applied to large data sets with N's in the 10,000's. Thus, the Rasch model was a reasonable candidate because of its simplicity. The proposed algorithm does indeed provide results relating well to LOGIST, and the writers believe it has merit.

There are additional matters yet to be considered. For example, fit of the items to the model was only cursorily considered by the investigators; it is felt that some simulation studies should be carried

out; that additional attention needs to be directed towards the divergence of estimates for high ability levels; and that thought should be given to the philosophical and psychological aspects or scoring when omitted responses are present.

REFERENCES

Lord, F.M. and Novice, M. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

Wingersky, M.S. and Lord, F.M. "A Computer Program for Estimating Ability and Item Characteristic Curve Parameters When There are Omitted Responses." Princeton, N.J.: Research Memorandum No. 73-2, Educational Testing Service, 1973.

Wood, R.L. and Lord, F.M. "A User's Guide to LOGIST." Princeton, N.J.: Research Memorandum No. 76-4, Educational Testing Service, 1976.

Wright, B.J. and Mead, R. "BICAL: Calibrating Items and Scales with the Rasch Model." Research Memorandum No. 23A, Chicago: Statistical Laboratory, Dept. of Education, University of Chicago, 1978.

TABLE 1

Percent Omissions at End of Test	Difference in Ability Estimates for 70% of Subjects
0 - 20	.26 or less
20 - 80	.73 or less
90 - 100	.90 or less

TABLE 2

In this and succeeding tables, the following definitions hold:

$\hat{\theta}_R$ - ability estimates from BICAL

$\hat{\theta}_{1A}$ - ability estimates based on one parameter algorithm

$\hat{\theta}_{2A}$ - two parameter ability estimates

θ_T - three parameter LOGIST ability estimates

CORRELATION BETWEEN $(\hat{\theta}_{1A} - \hat{\theta}_R)$ AND:

(1)	Number of Omitted Items	.76
(2)	Proportion of Omissions at End of Test	.47
(3)	Mean Difficulty of Omitted Items	-.30

TABLE 3

CORRELATIONS AMONG ABILITY ESTIMATES

Above Diagonal: Entire sample (N=1000)

Below Diagonal: (1) 90% or more omissions at end of test
(2) 20% or fewer omissions at end of test

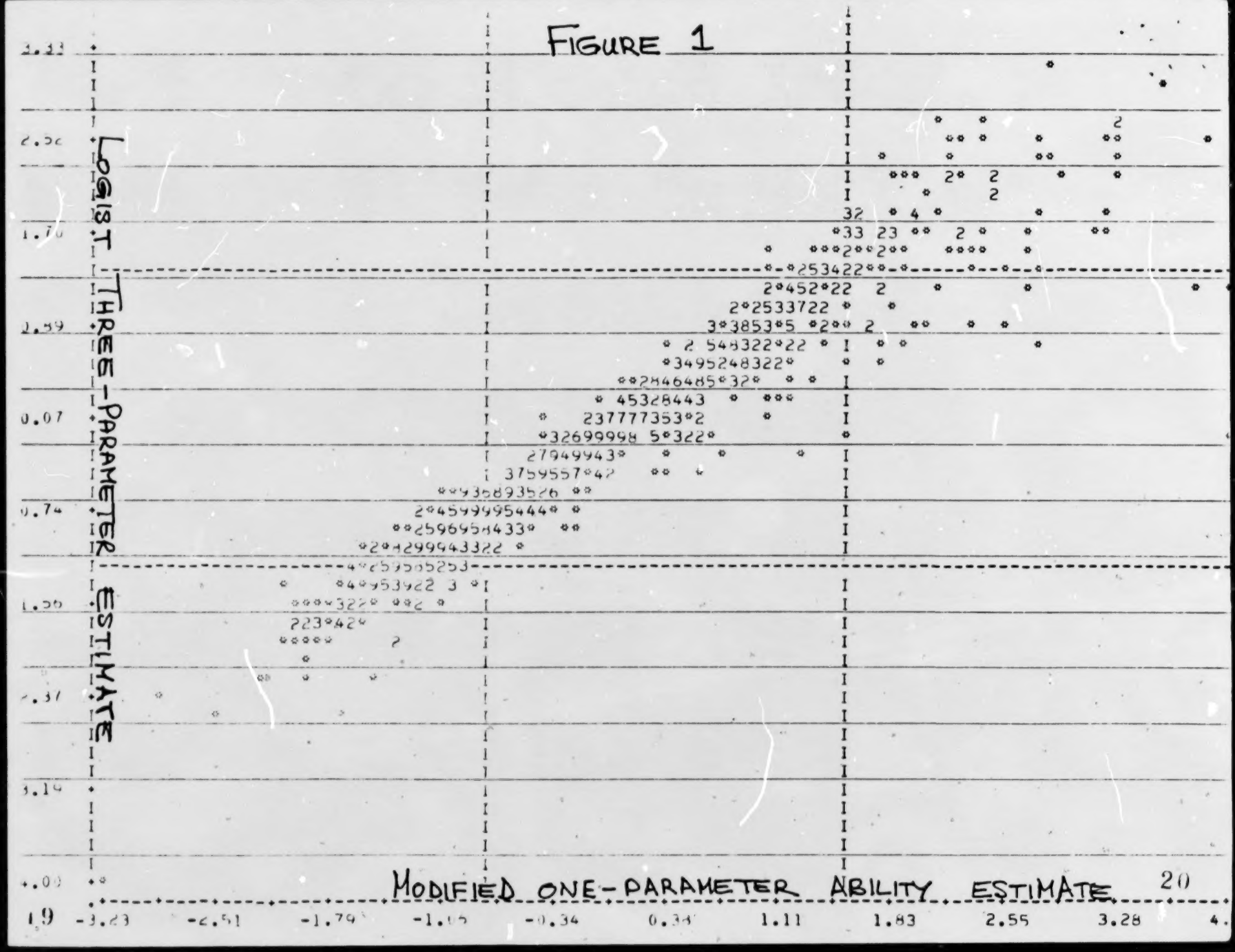
	$\hat{\theta}_R$	$\hat{\theta}_{1A}$	$\hat{\theta}_{2A}$	$\hat{\theta}_T$	CTB
$\hat{\theta}_R$	1	.92	.92	.91	.99
$\hat{\theta}_{1A}$	.89 .97	1	.96	.96	.92
$\hat{\theta}_{2A}$	.90 .99	.98 .96	1	.99	.91
$\hat{\theta}_T$	.89 .99	.98 .96	.99 .99	1	.99

TABLE 4

CORRELATION BETWEEN $\hat{\theta}_{IA}$ AND $\hat{\theta}_{T}$ AS A FUNCTION OF NUMBER OF OMISSIONS

Number of Omissions	Correlation
3 - 7	.98 (N=554)
8 - 17	.94 (N=327)
18 - 22	.87 (N=119)

FIGURE 1


LOGIT
THREE-PARAMETER ESTIMATE
MODIFIED ONE-PARAMETER ABILITY ESTIMATE
-3.23
-2.51
-1.79
-1.07
-0.34
0.38
1.11
1.83
2.55
3.28
4.